

```

1 CGCGGTGAAC AGCACGAGGG GGAGCGCTTG GCACGCGGAG CCAGAGCCGG
51 AGCTGCAGCC GCAGCGGGAG CCGGGGGAGC TCAGGGGCCG CAGGAGCCGG
101 GCCGGAGTGA GCGCACCTCG CGGGGCCCTC GGGGCAGGTG GGTGAGCGCC
151 ACCCGGAGTC CCGCGCGCAA CTTTCAGGGC GCACTCGGCG GGGCGGCTGC
201 GCGGCTGCCG GGACTCGGCG CGGGACTGCA TGGAGGCCAA GGAGAAGCAG
251 CATCTGTTGG ACGCCAGGCC GGCAATCCGG TCATACACGG GATCTCTGTG
301 GCAGGAAGGG GCTGGCTGGA TTCCTCTGCC CCGACCTGGC CTGGACTTGC
351 AGGCCATTGA GCTGGCTGCC CAGAGCAACC ATCACTGCCA TGCTCAGAAG
401 GGTCTTGACA GTCAGTGTGA CCCCAAGAAG GGAAGGCCCC AGCGCCAGCT
451 GTATGTAGCC TCTGCCATCT GCCTGTTGTT CATGATCGGA GAAGTCGTTG
501 GTGGGTACCT GGCACACAGC TTGGCTGTCA TGAAGTACGC AGCACACCTG
551 CTCACTGACT TTGCCAGCAT GCTCATCAGC CTCTTCTCCC TCTGGATGTC
601 CTCCCGGCCA GCCACCAAGA CCATGAACTT TGGCTGGCAG AGAGCTGAGA
651 TCTTGGGAGC CCTGGTCTCT GTACTGTCCA TCTGGGTCTG GACGGGGGTA
701 CTGGTGTACC TGGCTGTGGA GCGGCTGATC TCTGGGGACT ATGAAATTGA
751 CGGGGGGACC ATGCTGATCA CGTCGGGCTG CGCTGTGGCT GTGAACATCA
801 TAATGGGGTT GACCCTTCAC CAGTCTGGCC ATGGGCACAG CCACGGCACC
851 ACCAACCAGC AGGAGGAGAA CCCAGCGTC CGAGCTGCCT TCATCCATGT
901 GATCGGCGAC TTTATGCAGA GCATGGGTGT CCTAGTGGCA GCCTATATTT
951 TATACTTCAA GCCAGAATAC AAGTATGTAG ACCCATCTG CACCTTCGTC
1001 TTCTCCATCC TGGTCTGGG GACAACCTTG ACCATCCTGA GAGATGTGAT
1051 CCTGGTGTG ATGGAAGGGA CCCCCAAGGG CGTTGACTTC ACAGCTGTTT
1101 GTGATCTGCT GCTGTCGGTG GAGGGGGTAG AAGCCCTGCA CAGCCTGCAT
1151 ATCTGGGCAC TGACGGTGGC CCAGCCTGTT CTGTCTGTCC ACATCGCCAT
1201 TGCTCAGAA ACAGACGCCC AGGCTGTGCT GAAGACAGCC AGCAGCCGCC
1251 TCCAAGGGAA GTTCCACTTC CACACCGTGA CCATCCAGAT CGAGGACTAC
1301 TCGGAGGACA TGAAGGACTG TCAGGCATGC CAGGGCCCCT CAGACTGACT
1351 GCTCAGCCAG GCACCAACTG GGCATGAAC AGGACCTGCA GGTGGCTGGA
1401 CTGAGTGTCC CCCAGGCCCA GCCAGGACTT TGCCTACCCC AGCTGTGTTA
1451 TAAACCAGGT CCCCTCCTG ACCTCTGCCC CACTCCAGGA ATGGAGCTCT
1501 TCCAGCCTC CCATCTGACT ACAGCCAGGG TGGGGACTCA GCGGGTATAA
1551 AGCTAGTGTG ACCCTGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
1601 AAAAAAAAAA AAATGTT

```

FEATURES:

Start: 230

Stop: 1346

HOMOLOGOUS PROTEINS:**Top BLAST Hits:**

| | Score | E |
|--|-------|-------|
| gi 6981714 ref NP_037022.1 zink transporter 2 >gi 8134839 sp Q... | 583 | e-166 |
| gi 8134844 sp P97441 ZNT3_MOUSE ZINC TRANSPORTER 3 (ZNT-3) >gi ... | 390 | e-107 |
| gi 4508043 ref NP_003450.1 zinc transporter 3 >gi 8134847 sp Q... | 379 | e-104 |
| gi 6755552 ref NP_035903.1 solute carrier family 30 (zinc tran... | 349 | 3e-95 |
| gi 7507918 pir T24963 hypothetical protein T18D3.3 - Caenorhab... | 327 | 1e-88 |
| gi 7297930 gb AAF53175.1 (AE003635) CG17215 gene product [Dros... | 310 | 1e-83 |
| gi 7019533 ref NP_037441.1 solute carrier family 30 (zinc tran... | 279 | 4e-74 |
| gi 7106411 ref NP_035904.1 solute carrier family 30 (zinc tran... | 275 | 7e-73 |
| gi 7509701 pir T26757 hypothetical protein Y39E4A.2b - Caenorh... | 274 | 9e-73 |
| gi 7509700 pir T26756 hypothetical protein Y39E4A.2a - Caenorh... | 274 | 1e-72 |
| gi 8134837 sp O55174 ZNT4_RAT ZINC TRANSPORTER 4 (ZNT-4) (DRI 2... | 273 | 3e-72 |
| gi 7630076 emb CAB88298.1 (AL353032) zinc transporter-like pro... | 246 | 2e-64 |
| gi 7485763 pir T02681 hypothetical protein F19D11.8 - Arabidop... | 238 | 7e-62 |

EST:

| | Score | E |
|--|-------|-------|
| gi 5661927 gb AI925963.1 AI925963 wh12g11.x1 NCI_CGAP_Kid11 Hom... | 143 | 1e-31 |
| gi 5434480 emb AL046404.1 AL046404 DKFZp434N107_r1 434 (synonym... | 62 | 4e-07 |

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

Expression information from BLAST EST hits:

gi|5661927|gb|AI925963.1 Human kidney
gi|5434480|emb|AL046404.1 Human testis

Expression information from PCR-based tissue screening panels

Human Kidney
Human Heart
Human Placenta
Human Small intestine
Human Liver

```

1  MEAKEKQHLL DARPAIRSYT GSLWQEGAGW IPLPRPGLDL QAIELAAQSN
51 HHCHAQKGPD SHCDPKKGKA QRQLYVASAI CLLFMIGEVV GGYLAHSLAV
101 MTDAAHLLTD FASMLISLFS LWMSSRPATK TMNFGWQRAE ILGALVSVLS
151 IWVVTGVLVY LAVERLISGD YEIDGGTMLI TSGCAVAVNI IMGLTLHQSG
201 HGHSHGTTNQ QEENPSVRAA FIHVIGDFMQ SMGVLVAAYI LYFKPEYKYV
251 DPICTFVFSI LVLGTTLTIL RDVILVLMEG TPKGVDFSTAV RDLLLSVEGV
301 EALHSLHIWA LTVAQPVLVS HIAIAQNTDA QAVLKTASSR LQGKFHFHTV
351 TIQIEDYSED MKDCQACQGP SD

```

FEATURES:

Functional domains and key regions:

| <u>InterPro</u> | <u>Results of</u> <u>FPrintScan</u> against PRINTS | <u>Results of</u> <u>HMMPfam</u> against PFAM-A | <u>Results of</u> <u>PPsearch</u> against PROSITE | <u>Results of ProfileScan</u> against <u>PROSITE</u> profiles |
|-----------------------------------|--|---|---|---|
| IPR002524 Cation efflux family | | PF01545 [127-358] | | |

Membrane spanning structure and domains:

| Helix | Begin | End | Score | Certainty |
|-------|-------|-----|-------|-----------|
| 1 | 74 | 94 | 1.565 | Certain |
| 2 | 106 | 126 | 1.435 | Certain |
| 3 | 141 | 161 | 2.047 | Certain |
| 4 | 176 | 196 | 1.693 | Certain |
| 5 | 224 | 244 | 1.165 | Certain |
| 6 | 250 | 270 | 1.954 | Certain |
| 7 | 305 | 325 | 1.184 | Certain |

BLAST Alignment to Top Hit:

```
>gi|6981714|ref|NP_037022.1| zink transporter 2
>gi|8134839|sp|Q62941|ZNT2_RAT ZINC TRANSPORTER 2
(ZNT-2) >gi|2144024|pir||S70632 zinc transporter ZnT-2 -
rat >gi|1256378|gb|AAB02775.1| (U50927) zinc transporter
ZnT-2 [Rattus norvegicus]
Length = 359
```

Score = 583 bits (1487), Expect = e-166

Identities = 290/362 (80%), Positives = 325/362 (89%), Gaps = 8/362 (2%)

```
Query: 15 AIRSYTGSLWQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCDPKKGKAQRQL 74
A RS+ G+LW+ A IP ++L ++ELA QSNH+CHAQK SH + +K +A+R+L
Sbjct: 2 ASRSFFGALWKSEASRIP----PVNLPSVELAVQSNHYCHAQKDSGSHPNSEKQRRARRKL 57
```

```
Query: 75 YVASAICLLFMIGEIVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNF 134
YVASAICL+FMIGE++GGYLA SLA+MTDAAHLLTDFASMLISLFSLW+SSRPATKTMNF
Sbjct: 58 YVASAICLVFMIGEIIIGGYLAQSLAIMTDAHLLTDFASMLISLFSLWVSSRPATKTMNF 117
```

```
Query: 135 GWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVAVNIIMGL 194
GWQRAEILGAL+SVLSIWVVTGVLVYLAV+RLISGDYEI G TMLITSGCAVAVNIIMGL
Sbjct: 118 GWQRAEILGALLSVLSIWVVTGVLVYLAVQRLISGDYEIKGDTMLITSGCAVAVNIIMGL 177
```

```
Query: 195 TLHQSGHGHSHG----TTNQEEENPSVRAAFIHVIGDFMQSMGVLVAAYIYFKPEYKYV 250
LHQSGHGHSHG ++QQ++NPSVRAAFIHV+GD +QS+GVLVAAYI+YFKPEYKYV
Sbjct: 178 ALHQSGHGHSHGHSHEDSSQQQQNPSVRAAFIHVVGDLQSVGVLVAAYIIYFKPEYKYV 237
```

```
Query: 251 DPICTFVFSILVLGTTTLTILRDVILVLMEGTPKGVDFTAVRDL LLSVEGVEALHSLHIWA 310
DPICTF+FSILVLGTTTLTILRDVILVLMEGTPKGVDFT V++LLLSV+GVEALHSLHIWA
Sbjct: 238 DPICTFLFSILVLGTTTLTILRDVILVLMEGTPKGVDFTTVKNLLLSVDGVEALHSLHIWA 297
```

```
Query: 311 LTVAQPVL SVHIAIAQN TDAQAVLKTASSRLQGKFHFTVTVTIQIEDYSEDMKDCQACQGP 370
LTVAQPVL SVHIAIAQN DAQAVLK A RLQGKF+FHT+TIQIE YSEDMK CQ CQGP
Sbjct: 298 LTVAQPVL SVHIAIAQNVDAQAVLKVARDR LQGFNFHTMTIQIESYSEDMKSCQECQGP 357
```

```
Query: 371 SD 372
```

S+

```
Sbjct: 358 SE 359
```

Hmmer search results (Pfam):

| Model | Description | Score | E-value | N |
|---------|----------------------|-------|---------|---|
| PF01545 | Cation efflux family | 268.9 | 6.9e-77 | 1 |

Parsed for domains:

| Model | Domain | seq-f | seq-t | hmm-f | hmm-t | score | E-value |
|---------|--------|-------|-------|-------|----------|-------|---------|
| PF01545 | 1/1 | 127 | 358 | .. | 1 259 [] | 268.9 | 6.9e-77 |

1 CCTGCCACCA TGCCTGGCTA ATTTTCTTAT TTTTAGTAGA GACGAGGTTT
51 TGCCATGTTG ACCAGGCTGG TCTCGAACTC TTGACCTCAG GTGATCCGCC
101 TGCCTCAGCC TCCCAAAGTG CTGGGATTAT AGGCGTGAGC CGCCGCACCC
151 AGCCAACATT TTTTAAATAC TGAAGAGTAG AGGGAATAGT TATAGTGTAC
201 CCCATTTACC CATCACTCAG TTTCAACAGC TGGTGACATA TTTATTTCTT
251 CTATACCAGT ACCGTACTCT CCCCCTGGG ATTATTTTAA GGCAAAACCC
301 AGATGACATT TTATCCCTAA ATACTTTAGA TAAAGGTGTT CTTTGAAAAA
351 AATCATAACC TCAGGACCAG CCTGGCCAAC ATGGTGAAAC CCTGTCTGTA
401 CTAAAAATAC AAAAATTAGC TTGGCATGGT CGTGGGCACC TGTAATCCCA
451 GCTACTCAGG AAGCTGAGGC AGGAGAATCA CTTGAATCCG GGAAGCAGAG
501 ATTGCAGTGA GCTGAGATTG CAGTCGAGCC TGGGCGACAG AGACAGAAAT
551 GAAACTCTGT CTCAAAAACA AACAAACAAA AAAACCACTA TACATAAAAA
601 TGAACAATGA TGCCACAATA GCACCAGAGA ATTTTATAAA TACAGATTCC
651 CAGGCCCTGC CCCAGACCTA CTGAATCCTG GAAATATTCA GGCTCCACAC
701 CCAGAGATTG TGGTTTCGGT GGTCTGATGC AGGGACCTGT AACCTGCGTT
751 GTAACACCTT CTCCAGGTAA TGCTGAGCCT GCTGGTGCTC AGAGTAGACA
801 GACCTGGAGA AAACCAGGGT GTCTGAGGTT TTCCAGAAGA AAACCAGAGT
851 CCAGACAAGC AGAGAGGCAC TCAGTGAGGA CCCAAGCAGA GCGGGTGCAC
901 CTCACATCCT CACTCCTGGC ACCCGTCTC CTACAAGATG AGAGACTGAA
951 AGAGCCCCTT CCTGTCCCA GTGGTGTGGG CAAGAGGCCT GCACCTCTGA
1001 CTCTTGCTC TGTGAAAGGC CATACCCACC AAGCCTATGG TCCTAGCGAC
1051 AAAGGGTGCT GGGGAGACGA ATTGACCAGA CAGGGAGGTC TCCAGCAGCT
1101 TCTTTCTACA CAGAGGGCAC CTGTGAGAGG CCAGCGTGGG GGCCACAGGC
1151 TCCCCAATCC CCAAGAACCG CCAGGGAAGG AGGCTGCTTC AAGTGGGTGG
1201 GGCACCAAGC TGGCCAGGAA GGACAGGGCT TCTCCAGCG GTACCAACAC
1251 GGTGGCACCT CCGGCCTGCA TCTCCAGGC TTGCTTGTCG GGCTTCCTGG
1301 GGCTCCCAGG AGCCGCTGCG GGGGAGGGGA GAAGGGGTGG CAGCAGTGGC
1351 AGTGGTCGTC TCTGCTCCGA TGGTGACTGC CGATGACACT GTTCTCTGTG
1401 CGGGTGGAGA CAAAGCCGGC CACTCCAGAT TCTCCTGCGC GCAGGAGAGG
1451 AGGAGCTGGC GCTGCTTCAG TGGCGAGGAT GGGTCGATCA GTCCCAGCCG
1501 TCTCTAGGGA GAGACACTGC CCCAGCCTGA GGGCGGCGCA GCCCACCCCA
1551 CCCCAGGACC CTCTAGCAG GAGGACAGGA ACGCAAGCCG ACCTCGGGGG
1601 GTCTCCGGCC TGAGAGGGGA ACATGATCAA GCCCAGGGCA GCCGCCAGTC
1651 GGAGGGGGCA GACGCGGCCC CAGTAGCCTC TGGAGACCCT CTTCCGAGGC
1701 AAGGAGCCAC ATTCTGCCC GCGGGACCAC CAAAGCGGAT TTCTACAAAC
1751 TAAAGTCGAG AACTTTTCGG CCGCGAGGCG GCGCACCCCG CGCGGGAGAG
1801 GGGGCGCAGG CGTCACCCCG TCCTCACTCA GCAACACCCG GGCGCGCGC
1851 CGGGCGAAGG CTGGCAGACT TCTCGCGGCC GGCAGGTGGG CTCGCGGGG
1901 CCCATGGGCG CAGGCACAGG TGTGCGGGG CACAGCCGGG CCTTTTGCAG
1951 CCGGCGACCG CCCCCCTTC CCGCGGGGCT TTTGCACACG ACGCGCGGAC
2001 GGCAGCTTCA CACGGGTTGG CGAGGGCCGG ATAAAGCCGG CGGCCGCGGG
2051 GCGCAGCGGC TGACCGGAGA CACGGGAGCG CTTGGCACGC GGAGCCAGAG
2101 CCGGAGCTGC AGCCGCAGCG GGAGCCGGGG GAGCTCAGGG GCCGCAGGAG
2151 CCGGGCCAGG GTGAGCGCAC CTCGCGGGG CCTCGGGGCA GGTGGGTGAG
2201 CGCCACCCGG AGTCCGCGC GCAACTTTCA GGGCGCACTC GGCGGGGCGG
2251 CTGCGCGGCT GCCGGGACTC GGCGCGGGAC TGCATGGAGG CCAAGGAGAA
2301 GCAGCATCTG TTGGACGCCA GGCCGGCAAT CCGGTAAGGC GAGAGCCTGG
2351 GGACGGGCGG GAATTTTGGC GGCAGCCGAC GCACCTCCGC ATTTGCCATC
2401 CTAGCAGTGG TTGGTTTGG CTGCGGCTTG AGCTGGGACC ATGCAGGAGG
2451 GGTGGGTGTA GGTGAGGGAA CGAAGATAAT GGGCTGTGGC CCAAGGACCG
2501 TCTCTCCCTT GGGGCGCAGC CCAGATCCTG ACCCTGCCCC GCGGGGGGGC
2551 TGGGGCCGGG GGATTGGCGG TTCCCATGCC TGGAGTCCCG CCCC GCCCAG
2601 CACTTTGCCC CAGGCAGCCC CGCCCCGGGG AGGCCCTGTG TCCCAAGTGC
2651 GCTGGAGGGG GCCTGCTGTT CTCCAGAGC CTGCGCTCTG TTCTTTCCCC
2701 GCCTCCACT GGACAGCACG CCCCTTGGCC GGTTCACAGG GTCTTCACCT
2751 CCTCTGGCCT CTGAAGGGCC CCGGGCCCCA GGAATCCCAT TCCCCTATCA
2801 TCCCCGTCTG AATACAGGCT TCTCACCTCT GGTGTGTCGA GCTCGGAGCG
2851 TCCTTAGCTA TTTTCCAG TGGACACAAG GCTTCACAGA GAAATGGGAC
2901 TAGAGTCGGC CCTCCTTACC TCATCTCAGA GCTGAGCGTT CCCTCTCTTC
2951 CCCTCTGGCC AGGTCATACA CGGGATCTCT GTGGCAGGAA GGGGCTGGCT
3001 GGATTCCTCT GCCCCGACCT GGCTGGAAT TGCAGGCCAT TGAGCTGGCT
3051 GCCCAGAGCA ACCATCACTG CCATGCTCAG AAGGGTCTCT ACAGTCACTG
3101 TGACCCCAAG AAGGGGAAGG CCCAGCGCCA GCTGTATGTA GCCTCTGCCA

FIGURE 3, page 1 of 5

3151 TCTGCCTGTT GTTCATGATC GGAGAAGTCG TTGGTAAGCA CTTTTGGGCT
3201 AATTAAATGA AGTTGGTGCA TGGATAGACT GGATGTTCCC AGCAATACTG
3251 AACTAAAAAG CCCCAATTAC TGAACACACA CTACAGTAAG CCTTTATATA
3301 CACATTATCT GATGCAGGTC TAACAACAAC CTGTGTTCTC ACATGGGTGA
3351 CGTTATTCTC CTTACTTTAC AGATGATGAA ACTGAGGCAC GTTCAGGTGA
3401 AGTAACTTGC CAAAGATCAC GCACAACCTCG TAGCTAAGGG AAGGCCTGAA
3451 TTCCTAGAAG GGAAGAGCAT TTACTGAGTA TCTGCTATGT TTTCCAGTCC
3501 CTATTTGAAC TTGATGTGCA CATTCAACCT CTAAGTAGAT ATTGGTGGCC
3551 CATTTCACAG AGAGTGGAAG TTGAGGCTCA GAGAGAGTAG GTCACTTGTC
3601 ACGGTGGTAC AGCTCATGGG TAGAGAGATC TTGAGCCCAG AATGTCCTCT
3651 TCCAGAGCCT TGCGTCTCCC TGCTGCACAC ACAGTCTTGG GAGCCAGCTC
3701 TCTGGGGGAG CTGATAAGGA CCCTCCACCC TGCAGGTGGG TACCTGGCAC
3751 ACAGCTTGGC TGTCATGACT GACGCAGCAC ACCTGCTCAC TGACTTTGCC
3801 AGCATGCTCA TCAGCCTCTT CTCCCTCTGG ATGTCCTCCC GGCCAGCCAC
3851 CAAGACCATG AACTTTGGCT GGCAGAGAGC TGGTGAGGAT CGCGGTTTGG
3901 CTGGAGATGG GGTGAGAGA GAGGGTGGGT TAGAACAGGG GTTCTTAGGT
3951 GGCTGTAATG GGTGGATCCC CCCTTCCTCC CCTGAGTGAG GCCAGGAGGG
4001 TGATCTGGAT GGGGAAGAG GATGTCAACC ATGGCCTCTG TCCTCTGGGA
4051 AATCCTAGTC TGATGGGGGA GCCCTGGTCC CAGTCATCCA GGAGCTCTCA
4101 GTCTGCAGGG AAGCAAAGTT GACCTTCCTA AGAAGTGCAG TAGCCAAGCT
4151 TCAAGAACAA ATGACAATGG CATTAACT GCACATAACT CTGTGGATCA
4201 GCTCTGGGGG GAGGGGAAGG CCAGCAAAGG CTGCTGGAAG ATATAGGCTT
4251 TAACCTCTCT TCCGTTCCAC CTGGACTGCA TCGTCACCTT CCTCTTTGTG
4301 GGAGATGGCC CAGCCTGTCT TCCCCAGAAG CCTCAGTTTA CTAGCTGAAC
4351 AAAAGGCACA TACTTTAATA AGTCAGCTTC TTTACATGTA CAACCAAAAA
4401 GGTGGACTCA GATGATGACT TATTAGTTCC TTCTAGCTCT TACATTCTTA
4451 GATATATGAG GGGTGGGGTA AGGGGCAGCC ATACCGGCCC ATACCTTCAG
4501 CAGAGCCCAG TGGGACCCAG GCCCTGACTT TGTGTGGGAG GTGGGTGGGG
4551 AGGATCCTGA AGGAAGGGGG AAGACTCCTT AGCTCCAGGC CCATGCCAAG
4601 GTGGGTGTTG GGGTTGGGTT CTTCCTCAGA GATCTTGGGA GCCCTGGTCT
4651 CTGTAGCTTC CATCTGGGTC GTGACGGGGG TACTGGTGTA CCTGGCTGTG
4701 GAGCGGCTGA TCTCTGGGGA CTATGAAATT GACGGGGGGA CCATGCTGAT
4751 CACGTCGGGC TGCCTGTGTC CTGTGAACAT CATGTGAGTG GGGCCCCAGT
4801 TTCCCTCGTC TCCCTCCTC CTCCCGCTC TCACACCCAC ACCTATGTCT
4851 GCTTTCGGGA AAGAGACTGT GCCACTTTCC AGCATACGCT ACAGGGACAG
4901 AACTTCCCTA ATGGTCTGAG CTCTGGCACC TGGAACACCT GGGTCCCTACC
4951 TTAGGCCTAG GCCAAGAACA CTGGGAGCTG TAAATCGGAG TCTTCATCCA
5001 CTGTACCCAC TCCCTGATAC ATGTCAGGGA CTAGCCTTGG TGCGCTCATA
5051 CCTGAAGTGG GCGGGGAAGA GGCCAGTTGT TGCAGGAGTA GCTGTCCCTA
5101 GGGGCAGAAC CCAAGTCTGA AATTGGTCTC AGTTAGAGAC AATGGGTGTC
5151 TCTTTCGGGG TCTTTGTTCA GAGGCCTCAG TTTCCCCATC TGTGACATGA
5201 TGGAGTGAAC TGACAGTGAC CTCCCTAATG CCCTCCTGCT CTGAGATTTG
5251 ACATGTGCGC ATTGTTGTGC CCAGGCTCAG CCTGGCATTG GCGCTGGGCC
5301 CTATCTCTCA TGGCTGTCTG AACCAAGGCC ACGTGGGTTG GACTTCTCAC
5351 ATGGCCAAAG AGATCACAAG GTTTAGGGGC TTGAGATTTT TGCCCTACAA
5401 GTTGGCTAGT CCTAATAGGT GACCTCCATC TGCAGCTCA GTGAGCCCTT
5451 GGCTTTGTCT CCACTTCCAT AGAATGGGGT TGACCTTCA CCAGTCTGGC
5501 CATGGGCACA GCCACGGCAC CACCAACCAG CAGGAGGAGA ACCCCAGCGT
5551 CCGAGCTGCC TTCATCCATG TGATCGGCGA CTTTATGCAG AGCATGGGTG
5601 TCCTAGTGCG AGCCTATATT TTATACTTCA AGGTCAGAGC TGGGACACAG
5651 TGTGGTGGGG GTGGCAGGGG AGTGTAGACC ACCTGAGTAT ACTCTCTACC
5701 GGGGTTTCTT TTCAGATTCT AGCTCCCTCC CAGTTCTAGG GAAAAGGGTG
5751 GGGAGAGGAA AGGAACATTT ATCCAATACC TACCAAGTGT CAGCACTTCT
5801 GATCCTCACA ACAACCTGAA GGGTAGGTGG TAGTGTTTTT TGTAGCTCAG
5851 AAAGGTTTCTG TGAAGTGCAC AGTGTACAC AGCCGGTAAA GCATAGAGCC
5901 AGATTCAAGC CTACGACTGT GTGTTGTCAA ACCCTGGGCA ATGCCCATCA
5951 CATAGAGGCA GGGAGCTGTA GTGGAAAGAG GCAGGCATTT GCTCTGAAGC
6001 TTGGCTCTCC TCCTTGCTAG CCATGTGACA TTGGATGAGT TTGCTTGCTC
6051 TAATGGAGCC TCAATTTCCC CATCTGTCAA ATGGGGACGG ATGGCGGATC
6101 AGATGGTATC TAAGATGGCT TTTTGCTCTG TCCATGTCTC AGCTCCTTGA
6151 GAAGGAGGGG TGGGAAGGGA CTGCCTTATT CTGAACTGTG GTCTGTCTCT
6201 TCTGCTCTTG CAGATGTGAA ATAAAGAGCA GAAAAGTGGG AGGCAGGGCC
6251 AGGGCGAGGC TCATGCCAC CCAGCAGAGA GAGCACCTCT CCCAGCAGT

FIGURE 3, page 2 of 5

6301 GCTGGGTGGG AGGGGAGAAG GGAAGCTGAG GTGTTAGATG GTGAACTCCA
6351 GGTCTGCCTT CCTGTCTTCC TGCAGCCAGA ATACAAGTAT GTAGACCCCA
6401 TCTGCACCTT CGTCTTCTCC ATCCTGGTCC TGGGGACAAC CTTGACCATC
6451 CTGAGAGATG TGATCCTGGT GTTGATGGAA GGTAACCTGG GCTTTGTGGC
6501 TCCTTTTTTG CTCTTGGCTC TCAAGCGCTA ATCAGCTCAA ATAGGGTATG
6551 TGTGTGTCTG GGGCATCCTA GCACATGGGC GGGGAGCCAG GATCCGGAGC
6601 CCCGGCATAG GCTGGAAAAC CTCCTGGGGC CCCTGGGCTG ATCTTGACAT
6651 AGAGCCTGGG CTTTCAGGTG TGGCAGTTCC TGGAAACCGTC CCCCAGCCCCG
6701 AGTCTTCCCT TCCCCCTACC CCTAAGGGTG CCTCCTCTGC CTAGTCAGGT
6751 GGCTTCTGGG GGACATCTGT AGCATCTGGA GCTCTCCAGC CCTCCCCTAT
6801 ACACTTCCCC AGGCTCTGGC TGCCTTCTCT CAGGAAGAGA GAGGGGGTGA
6851 GGATTATGCT TCTCATTGCA CAGAGGGGCA GACTGAGGCT CAGAGAAGGA
6901 CAGTCAGCCT TGGACAAAGC TACTGAATCC ACTGCAGCGC AGGCCTTTCC
6951 TACATCTCAG GGACCAAACA ATGCCACACC CTGTGGGGAC ATGGCTGTGC
7001 TTTGTGGGGT TGGAGAACGG TCAGTGGTGG AGAATGATCT GGTCTGCCCT
7051 GAATTACCTT TTTTTTTTCT TTTTCTTTT TTTTGAAACA GGGTCTTGCT
7101 CTGTCATCCA AGCTAGAGTG CAGTGGTGCC ACCAAGGCTC ACCGCAGCCT
7151 TGACCTCCTA GGCTCAAGTA ATCCTCCTGC CTCAGCCTCC CAAGTAGCTG
7201 GGACCACAGG CGCATGCCAC CATGTCTGGC TAACTTTTAA ATGTTTGTAG
7251 AGATGGGGGG GGGGGGGTCT CACTATGTTG CCCTGGCTGG TCTCGAACTT
7301 TTGGGCTCAA GCAATCATCT CACTTCGGCC TCTCAAAGTG CTGGAGTTAC
7351 AGATGTGAGC CACCACACCT GGCCCTGCAC CTTGCTTTTC TTATGCTCTA
7401 GGCTTGGGTG CTGGGCCANN NNNNNNTTT CCTTCAAAT ATATTTATTG
7451 GCCAGGTGCG GTGGCTCACA CCTGTAATCC CAGCACTGTG GGAGGCCGAG
7501 GCAGGCAGAT TACCTGAGGT CAGGAGTTCA AGACCAGCCT GGCTAACACA
7551 GTGAAACCCT GTCTCTACCA AAAATACAAA AATTAGCCGG GCGTGGTGGC
7601 ATGCGCCTGT AGTCCCAGCT ACTCAGGAGG CTGAGGCAGG AATAATTGCT
7651 TGAACCAGGG AGGCAGAGGT TGCAGTGAGC CAAGATCACG CCACTGCACT
7701 CCAGCCTGGG TGACAGAGCA AGATTCCGTC TCAAAAAAAA CCAAAAAATA
7751 TATTTATTGA GCACCTACTA TGGAGTAGGT GCTGTTTTAG GCACCAAGGA
7801 TACTGTGGTA ATCAAAGGAG ACTGTCTGCT CCTCATGGAG TGTCCATTTT
7851 AGAGGGAGAA ACTGACAATA AGTACATTCA TAAATAATTT CAGTGTTAAG
7901 AGTGGAGAGG AAATACAACA GAGTGATAGG GCAGAGACCT TGGGAGGTGA
7951 AGGCAGCCTC AGACCTGCAG GCCAAAGAGG TCTTCTTTGA GGGGATGACA
8001 CCTGAGGATC AGGAGCCAGC CCTGCACCAA TGGGCAGGCG TGGGAGGGGT
8051 AGTTTCCTTT AGTTTCCCCT GTCCCTTGCC GTCCCTCAGG ACCCCCAAGG
8101 GCCTTGACTT CACAGCTGTT CGTGATCTGC TGCTGTCGGT GGAGGGGGTA
8151 GAAGCCCTGC ACAGCCTGCA TATCTGGGCA CTGACGGTGG CCCAGCCTGT
8201 TCTGTCTGTC CACATCGCCA TTGGTGAGTG CTTGGGACAC TCAGGGTGGG
8251 GTGGGAGACA GGCAGCCAAA GGCCTAGTGC CATCCCCAAC GGGTCCAGGT
8301 GACCCAGAT GCTCACAGTG CCCATGCATC AAGCCCAGCC TCATGCTGAG
8351 TACTTGATAC GCATTATTCC ATCTGATCAG CACAATCTCA TTTATCCATG
8401 AAGAACTGAT GGCTGGGGTT GGGTGGTAAA GTTACTTGCC CAGGCTTTTA
8451 CAGCTAGTGA ATGGCAGTAG GTGGCAGATT CCTGGCCTTA AGGCCAGTGC
8501 TTTACAGCT CTTTCAGGCA TGAGCCAGGT CTGGGCTGGG AGGCTACCTG
8551 GCAGAGGAAT GGAATCTGGG GGCTTCTCCA TGTTCATGGT CCCCCATCCT
8601 GTTCTGCTGG GGATGGGGTA TGAGATTTGG GCTCCTGATG GTTCCAAAGG
8651 GCCAGAGTAA ATGGCTTCCC CCGCTGTGTC CTCTCGGCCC CCAGCTCAGA
8701 ATACAGACGC CCAGGCTGTG CTGAAGACAG CCAGCAGCCG CCTCCAAGGG
8751 AAGTTCCACT TCCACACCGT GACCATCCAG ATCGAGGACT ACTCGGAGGA
8801 CATGAAGGAC TGTCAAGCAT GCCAGGGCCC CTCAGACTGA CTGCTCAGCC
8851 AGGCACCAAC TGGGGCATGA ACAGGACCTG CAGGTGGCTG GACTGAGTGT
8901 CCCCAGGCC CAGCCAGGAC TTTGCCTACC CCAGCTGTGT TGTAAACCAG
8951 GTCCCCCTCC TGACCTCTGC CCCACTCCAG GAATGGAGCT CTTCCCAGCC
9001 TCCCATCTGA CTACAGCCAG GGTGGGGACT CAGCGGGTAT AAAGCTAGTG
9051 TGACCCTGCT CTTCAGCTC CTGGGCCAGC TCTGGAAGGG CTGTATTGG
9101 GCCTAATCCT CAGCAAATGT TCTACCACTC GCAGGGGCAA AGTGGTGAG
9151 CCACGGGACG TCCAAGGGTA GGCTGGCCCC AGCGCGCCA TACTGCCTGC
9201 CTCATGCCCC ATTCTCAGCC TGGCTGGCCT TTGCCTTTAT GAATCTGAGC
9251 CCCTCCATCT GCCTATAGCA ATAGGCACGG GGGTGAGGAC CCTCACACTC
9301 TCATTTGAGC CTCCCTGAGG CAGGGAGCCA GGAGGCACCT GAGGCCTATC
9351 TGTGCCCTTAG TCACTTCAGC TATGAGCCAA ATGTTCCCTT TCCTGGAGGG
9401 GAGAGGCTTC TTACTAGGTA AGAGACAGGT TTCCTCTTTC CTTATTTCTT

FIGURE 3, page 3 of 5

```

9451 CAGCTGTGCC AACACAAAAA ACAACTTTGG CACAGGTGGT GGGCAGGGGG
9501 TAGAGAGATT TCAGCTTGGG TTCTGCACTA ACAGCCTCCA AGCCCCCTGG
9551 CACTTCTGTT GCCCTGAGAG TGTCCCAGGG GATTGAGAGT CTCCAGAAAG
9601 ATATGGCTGG GCCAACTCTG TTGCCTACCT GGCCTGACCC AGTCGGAGCC
9651 TGACATGGTG GAGGGAAAGG GAGACAAGTG GGGCTGCACT CGGTCCAGAG
9701 GCCAGCTAGG AGGGAAACCG CAGCTTCCTG GGGCTTGTGT GTGAAGATTC
9751 CTGACTTAGG GGTGGCTTTT GTTTACAAGA TGCAAGAGGG GAAACCTGTC
9801 CCCGACTCAT CGAGACAACA TGCCCAGTTA TCAGGGAGTC CTGTGTCACA
9851 AGGTCTGTCT CTGCCATTGT AAGCAAGTGC CTTGGGCGAG CTGGCCTCTG
9901 CCCACAGTC TCATCTGTAC ACCGACAGGG TTGATGCCTC CCTCACAGGG
9951 TTGAGAACAA GAGCCAGTTG GCCAAGTACC TGTGGTTGTT GAAGATTGGT
10001 TACTTTTACC ATCCTGGGGA CAGGGAATC TGTGGCCCGA GGCTGCCTCA
10051 CTGAGGAGTC AGGTGGGCTT CCCAGCCTCC CCAGGGGCGAG TGCTGAGTTT
10101 GTCTTGACTG TTCTGGCCCA AGGTGGGAGG AGGTGGGTTT GGTCACTTGC
10151 CTCCCACCTT AAATCTCTGT CTTTCCATCT GTGAAATGAC CTCTTTGTGC
10201 CTTCCCAGCA CTGTCATCCT GATCGCCTGT GTTCTAGGTA GGTGGGTCTT
10251 TCAGCCCCCT CAGGTCTGTG AAAAGTCTGT GGAAAGCACT GGCCTGGAGA
10301 GGGGTGGGGG GTTGCTGGTG GGTGCTCCAT TCCACCACAA TCTCAGGGGA
10351 CTAACCTCC CCTACCCAAC TACCCACCC CCACCAAGC CATGGCAGGC
10401 CCCAGGAACT TGATCCTGGG CTTTGCCGTA TGCCAAGTCC TTACACCCCT
10451 CTAAGAGAC AGTCATTGGC TGGGCACGGT GGCTCATGCC TGCAATCCCA
10501 GCACCTTGGG AGGCTGAGGC AGGCAGATGA CTTGAGGCCA GGAGTTCGAG
10551 ACCAGCCTGG CCAATATGGC GAAACCTCAT TTCTACTAAA AATACAAAAA
10601 CTAACCAGGC GTGGTGGCTT GTGCCTGTAA TCCCAGCTAC TCGGGAGGCT
10651 GAGGCAGGAG AATCGCTTGA ACCGGGGAGG CAGAGGTTGC AGTGAGCTGA
10701 GATCACACCA CTGCACTCCA GCCTGGGCGA CAGAGCGAGA CTCCAGCTTA
10751 AAAAAAAAAA AAAAAAAAAA AAAAGGAGAC CATCACTGCT GTCCTGCATT
10801 CTTACAGATG AAAAAACAGG CTCAGAGGTT GAATCGTTTT CCTGAAGTCA
10851 GACAGCCAGT GCAGGCAGGT CTGGGATTTC TGCCTCATTT CGGTAGACCT
10901 TCCTCTACAG CAGGGTCTGG GGGCCTGTCG GTCTGCGCTG CCTGTTGGTA
10951 CAATACAAAC CCTGGGACC AGCAGTGCCC GGCCCATGGG TGAGGACATG
11001 CCAAGGCAGT TCAGTGTCTT GGGTGTACAC GCTGTGATTG GAAAGGTGCC
11051 TCTTTCACCT GGCTGGGCCT GGCATCCAGC GCCCTCCCCA CCCTGGGAAG
11101 G

```

FEATURES:

```

Start:      2284
Exon:       2284-2333
Intron:     2334-2962
Exon:       2963-3183
Intron:     3184-3735
Exon:       3736-3882
Intron:     3883-4629
Exon:       4630-4783
Intron:     4784-5472
Exon:       5473-5632
Intron:     5633-6375
Exon:       6376-6481
Intron:     6482-8088
Exon:       8089-8223
Intron:     8224-8694
Exon:       8695-8840
Stop:       8838

```


MAP POSITION:

Bac Accession #: AF153980.1

Human chromosome #: 1

ALLELIC VARIANTS (SNPs):

| Position | Major | Minor | Context |
|----------|-------|-------|--|
| 406 | t | c | gccatctgtgcaatgtgcaggtgccagttgcatgtgccatgctgtgtggc[t/c]gtgagcg tgtgtgtgttttcgtgtatgtgccatgcacgtatgtgctgcgtg |
| 852 | t | c | attcgtgtgcaccgtacacgtgtgccatttgcatgtatgctgtgcacgtg[t/c]ggcatgc atgtgtgccgtttgcatgccatgcatgtgttccttgctgtgc |
| 897 | g | a | acgtgcggcatgcatgtgtgccgtttgcatgccatgcatgtgttccttgc[g/a]tgtgccg tgctgttccatgcacgtgtgccgtgcatgtgccattcgcgtg |
| 1,433 | c | t | cgcaggagtggcccagacgggggtcagcagtgtagtacagctggccatg[c/t]ggttcct acagcttccaggcgtcagactctggcagaagggtgagaccct |
| 5,845 | t | c | ggacagagcccaggtgggggtggggcggggtccagcaccacggccagcac[t/c]gaccacc aggacccccggagccagcaccatggacagaaaactgcccaccag |
| 7,028 | g | a | cctgtcaattccggggggacggcagccagctcccgacaggggtcccc[g/a]ggtggcc ccccaccactgtataggccttggtcgtgggggtggggcggggg |

| POSITION | Allele 1 | Allele 2 | |
|----------|----------|----------|--------|
| 406 | t | c | Intron |
| 852 | t | c | Intron |
| 897 | g | a | Intron |
| 1,433 | c | t | Intron |
| 5,845 | t | c | Intron |
| 7,028 | g | a | Intron |